Title
Temporal Chunk Signal Reflecting Five Hierarchical Levels in Writing Sentences

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Van Genuchten, Erlijn
Cheng, P.C-H.

Publication Date
2010

Peer reviewed

Temporal Chunk Signal Reflecting Five Hierarchical Levels in Writing Sentences

Erlijn van Genuchten (e.genuchten@iwm-kmrc.de)
Knowledge Media Research Center, Konrad-Adenauer-Str. 40
72072 Tübingen, Germany

Peter C-H. Cheng (p.c.h.cheng@sussex.ac.uk)
School of Informatics, University of Sussex
Brighton, BN1 9QJ, UK

Abstract

Previous research on the temporal chunk signal has focused on the use of pauses in behaviour to probe chunk structures in working memory. On the basis of some of these studies, a hierarchical process model has been proposed, which consist of four hierarchical levels describing different kind of pauses. In this model, the lowest level consists of pauses between strokes within letters. On higher levels, there are pauses between letters, words, and phrases. Each level is associated with a larger amount of processing when retrieving these chunks from memory. The main aim of the present study is to test whether the temporal chunk signal can distinguish a fifth level, the sentence level. A secondary goal is to replicate the findings which were used to construct the hierarchical process model in a manner that overcomes some of the limitations of the earlier experiments.

Keywords: Temporal chunk signal, graphical protocol analysis, writing, working memory, sentences.

Introduction

Chunks have a fundamental role in information processing in the human cognitive architecture. Chunks are individual pieces of information grouped into larger units that increase our information retention (Caroll, 2004). It is widely accepted in Cognitive Science that the hierarchical storage and processing of chunks in working memory provides a fruitful basis for explaining a substantial range of the behavioural phenomena, such as recall from long-term and working memory and expert performance. This acceptance comes in part from various established methods that have been developed to infer the particular structure of chunks possessed by individuals from their behaviours. Such methods have made an important contribution to developing accounts of cognition in complex tasks. One method is computational modelling. Another method, on which this research focuses, provides information about the structure of chunks in memory by measuring pause lengths that occur in verbal or motor actions, the inter response latency. Chase and Simon's (1973) work on chess expertise and Reitman's (1976) work on Go experts are classic examples of the uses of pause lengths during the recall of a board containing chess pieces or Go discs. There are many other studies that have also exploited pause lengths in actions in order to define chunks. In one of the earliest studies using this approach, McLean and Greg (1967) studied the chunking of arbitrary letter sequences. Later on, Buschke (1976) examined the gradual acquisition of chunks comprising clusters of every day words originally presented in unstructured lists. Egan and Schwartz (1979) showed how electronics experts chunked components of electrical circuits in terms of their functioning. In all these studies, the duration of a pause preceding an action that generated an element of the domain (in these studies letters, words, and components respectively) is taken as an indication of whether the element is within a putative chunk or at the boundary between chunks. The term *temporal chunk signal*, TCS, is used in order to refer to the basic phenomenon that underpins the use of pause lengths to probe chunk structures. Typically, the TCS is often used in a binary fashion, which includes setting some threshold (e.g., 500 ms) as a criterion upon which to classify successively produced elements as intra-chunk if the pause length is less than the threshold, or as inter-chunk if the pause length is greater than the threshold.

The present experiment is a continuation of our studies on the nature and application of the TCS that is manifest in the process of writing and drawing, or more general graphical production. We call our general approach to using the TCS to study chunk related behaviour in writing and drawing tasks, *graphical protocol analysis*, GPA. A standard graphics tablet is used to record pen strokes. Pause lengths are computed by finding the difference in time between the lift of the pen from the tablet at the end of one stroke and the time at which the pen touches the tablet at the beginning of the stoke of interest: $\text{pause}_{\text{item}} = \text{time}_{\text{pen-down-current-item}} - \text{time}_{\text{pen-up-previous-item}}$. In our previous experiments, tasks with known hierarchical structures have been used, such as 'to be or not to be', so that each pause could be coded as intra-chunk or inter-chunk. For sentences and language-like stimuli identified pauses have included: intra-chunks pauses between strokes within a letter (e.g., second stroke of a 't', level 0 or L0); inter-chunk pauses between letters within a word (e.g., between 't' and 'o', L1); and inter-chunk pauses between words within a phrase (e.g., between 'to' and 'be', L2).

Our previous experiments have shown that the TCS is a richer source of information about chunk structure than just a binary signal. We consider that TCS within GPA has potential to be used as general technique for the study of various cognitive phenomena. In the domain of copying mathematical formulae, the TCS was able to distinguish participants who had four different levels of expertise in mathematics (Cheng & Rojas-Anaya, 2007). The TCS has also been used to distinguish between children with and without dyslexia (van Genuchten et al., submitted). Cheng, McFadzean, and Copeland (2001) have shown that the TCS reflects three distinct levels of processing when drawing geometric figures. That experiment showed the TCS to be present when drawings are made with pen on paper or with a mouse on a computer screen. Obaidellah and Cheng (2009) used the TCS to reveal the role of perceptual chunks and spatial schemes in different modes of drawing complex abstract diagrams. In Cheng & Rojas-Anaya (2005) participants wrote number sequences that had been memorized with different chunk structures. The TCS showed the existence of three levels corresponding to: pauses between strokes within a digit (e.g., second stroke in 'ヲ', L0, ≈90 ms); digit level chunks (e.g., between '1' and '2', L1, ≈280 ms); and digit group level chunks (e.g., between '111' and '222', L2, ≈440 ms). In Cheng & Rojas-Anaya (2006) the TCS again showed the existence of the same three levels of pauses when writing familiar and jumbled sentences (≈90, ≈270, ≈400 ms respectively). Finally, moving beyond three hierarchical levels, Cheng & Rojas-Anaya (2008) devised an artificial sentences copy task with four hierarchical levels (e.g., 'ITH* ITH* ITH*, ITH* ITH* ITH*') and found the same pattern of stroke, letter, word and phrase level pauses (≈90, ≈250, ≈440 and ≈600 ms respectively). On the basis of these studies, a hierarchical process model has been proposed (see Figure 1) to explain these patterns in terms of the depth first serial processing. In this model, the hierarchical structure of chunks corresponds to the amount of processing associated with different branch lengths of the hierarchy, with longer branches indicating longer pause lengths.

The main aim of the present experiment is to test whether the TCS can distinguish more than four hierarchical levels, by adding a fifth level of pauses between sentences, and thereby extend the previous findings. Is it simply the case that this fifth level in the chunk hierarchy will merely result in an additional amount of processing and a corresponding increment of pause duration? If so, will the increase in magnitude of the pause length be linear as is the case between the other levels? A secondary goal of the experiment is to replicate the previous findings in a manner that overcomes some of the limitations of the earlier experiments. In particular, the found significant effects existed at the level of individuals using pairwise comparisons of the pauses between levels, but typically involved relatively small numbers of participants. Hence, a subsidiary aim of the present experiment is to test whether

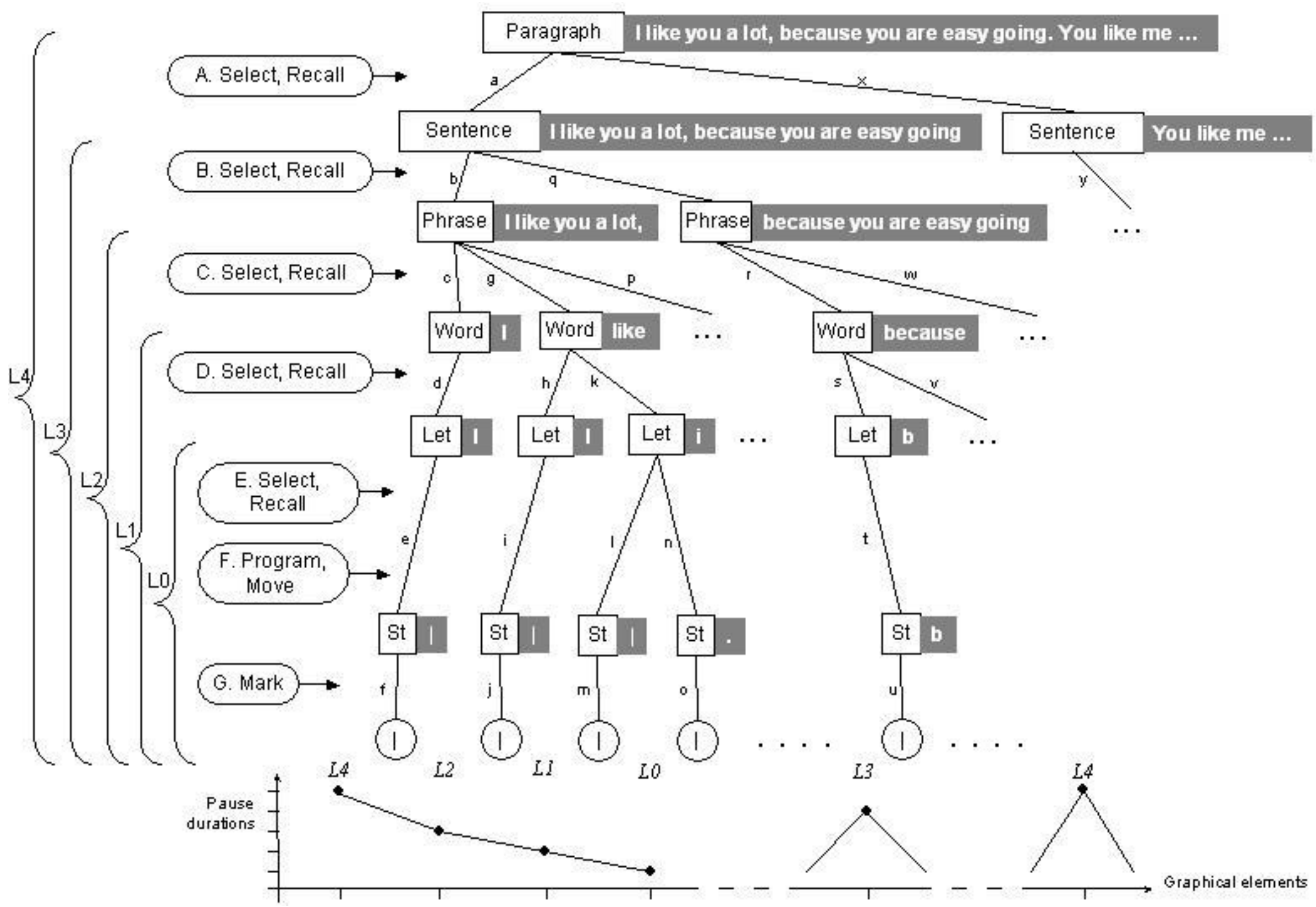

Figure 1: Hierarchical process model including the sentence level, reflecting relations between chunk structure, processing steps, and pause levels (adapted from Cheng & Rojas-Anaya, 2008).

the differences between the hierarchically levels is a robust effect by using multilevel analysis to simultaneously compare all the levels from the data of a large number of participants.

Method

Participants

Participants in this study were 32 adults, 19 female and 13 male, between 18 and 33 years old, working or studying at a large university in the UK. The participants (*M* = 22.99 years, *SD* = 3.98 years) were all native English speakers.

Measures and Materials

In order to answer the research question, pause lengths between sentences, phrases, words, letters, and strokes in paragraphs were compared. These five measures were calculated for each stimulus. The eight English sentence stimuli were specially written to obtain these five hierarchical levels. Each stimulus comprised of three or four sentences (L4), which were made up of two or three phrases (L3), which in turn were comprised of between 4 and 8 words (L2), which contained letters (L1) that may have required more than one stroke (L0) to write. This hierarchical structure was emphasized by including punctuation marks (periods between sentence and commas between phrases). Example stimuli are:

'You just signed up for a trip, from your favourite society, because you like visiting different places. You paid with some money, which you got from your mum, because you did shopping for her. You have never been to Holland, so you would like to visit Amsterdam, and have a great time.' (3 sentences, 3 phrases)

'We like swimming, in the pool next door. You like to cycle, to towns far away. They like to play football, on the top of the hill. As they play all day, they should eat enough.' (4 sentences, 2 phrases)

The median per level was calculated in order to reduce the influence of outliers, which could only occur in the direction of longer pause lengths, and which would consequently severely distort the mean, rendering it unsuitable as a measure in this study (Stavig & Gibbons, 1977).

All sentences were written on a piece of paper attached to a graphics tablet containing horizontal rows of rectangles. One letter had to be written in each rectangle (width: 6 mm, height: 8 mm), so that participants were encouraged to lift their pen from the paper and to put it down again for the next letter, and allowing the distance between each letter to be approximately equal (see Figure 2). The equal distance between rectangles rules out the possibility that differences in hand movements account for the different pause lengths. Therefore, pauses between the last letter of a line and the first of the next were ignored, because of increased hand movement.

Design and Procedure

The administering of the test had a duration of 45 to 75 minutes per participant. A quiet room was used to minimize disturbing background noises. The session began with an acclimatization period which allowed the participants to become familiar with writing on a tablet by having them write their names on the tablet. The actual experiment did not start until the participant was considered to have followed all instructions. Participants were asked not to write any punctuation in order to make sure that increased pause lengths were not due to writing an extra symbol. The task itself consisted of remembering and writing down the eight visually presented target stimuli.

All stimuli were presented in turn in random order. After presenting a stimulus, participants were allowed to apply any strategy and take as long as needed to rehearse the stimulus. When participants finished rehearsing, the experimenter tested recall accuracy by asking participants to recite the stimulus sentences without errors twice. Once this was accomplished, participants were allowed to start writing. A hash (#) had to be written at the beginning of each sentence to ensure that the writing process was well underway before the first letter was generated (Cheng & Rojas-Anaya, 2006).

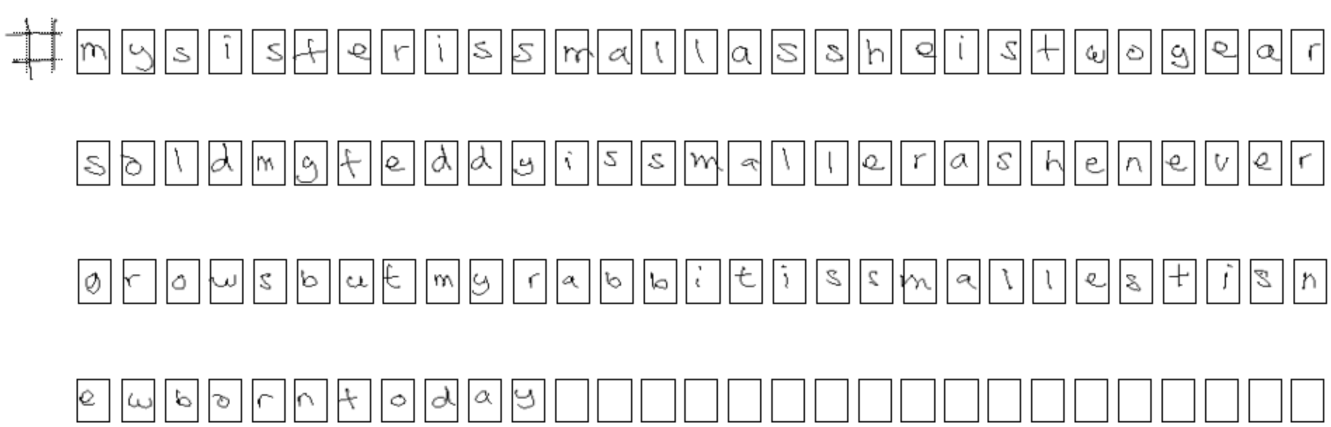

Figure 2: Example part of a written stimulus in equally spaced rectangles.

Table 1: Parameter estimates for multilevel models.

	Intercept only model Par.	SE	Model with predictors [a] Par.	SE	Robust standard errors [b] Par.	SE
Fixed effects						
Intercept	488	29.8				
Predictors:						
L0 (pauses between strokes)			90*	37.9	90*	3.1
L1 (pauses between letters)			273*	37.9	273*	13.6
L2 (pauses between words)			374*	37.9	374*	19.8
L3 (pauses between phrases)			567*	37.9	567*	41.3
L4 (pauses between sentences)			1134*	37.9	1134*	96.0
Random effects						
Predictor level variance	306818	7105.5	175177	7012.7	175177	7012.7
Participant level variance	20726	12282.6	24016	7101.0	24016	7101.0
Deviance	19846		19146		19146	

Notes: [a] The constant has been left out of the model with predictors, because a complete set of dummy variables was used.
[b] Robust standard errors were used, because the assumption of linearity of residuals was not met.
* $p < .001$.

Data and Analysis

A specially written program, TRACE, was used to record the writing actions and to extract the pen positions, times of points and pause lengths (Cheng & Rojas-Anaya, 2003). All files generated by TRACE were analysed using a specialized computer programme (Pause Level Extraction Tool, PLET, van Genuchten, 2009). Automatic detection of letters and automatic calculation of pause lengths had to be applied, because of the large amount of data (at least 1200 measures per participant). This involved identifying the horizontal position of strokes making up a letter within a rectangle and the horizontal separation between these strokes between rectangles. The written input was specified for each stimulus, so that errors (e.g., wrongly spelled or omitted words) were taken into account. Only in those cases where no automatic detection of letters was possible (e.g., when strokes of subsequent letters were written relatively close together), manual calculation of the pause lengths was applied.

To test whether there were differences between pause lengths of different pause types, a multilevel analysis was performed. On the lowest level, data of pause lengths of each stimulus for each pause type were used as predictors and were measured within participants. This data was gathered on the highest level, the participant. It is expected that the correlation between measures is higher within a participant than between participants. By performing a multilevel analysis, the dependency between data measures within individuals is controlled. Dummy variables were created for each of the five pause types.

Results

The comparison between the model-fit of the one-level and two-level intercept only models, indicates that there is significant variance at the participant level ($\chi^2(1) = 41.78$, $p < .01$). This means a two-level multilevel analysis is appropriate. Parameter estimates for the intercept only model and model with predictors are presented in Table 1.

A difference between the different pause types was expected, which was confirmed by the model with predictors. However, plots of standardized residuals against normal scores indicated that the assumption of linearity of residuals was not met. Therefore, robust standard errors were calculated using the Sandwich method (Hox, 2002). The resulting model indicates that a distinction can be made between pause lengths on the basis of pause type. Specifically, the regression coefficients show that pause lengths between sentences are longest, and that pause lengths become successively shorter when considering pauses between phrases, words, letters, and strokes (sentences: $B = 1134$, $SE = 96.0$, $p < .001$; phrases: $B = 567$, $SE = 41.3$, $p < .001$; words: $B = 374$, $SE = 19.8$, $p < .001$; letters: $B = 273$, $SE = 13.6$, $p < .001$; strokes: $B = 90$, $SE = 3.1$, $p < .001$). This means that pause lengths can be very well predicted when it is known which type of pause is concerned.

Discussion

One aim of the present experiment was to replicate the findings of previous experiments concerning the temporal chunk signal, TCS, using a more rigorous methodology. These earlier studies showed that the TCS reflects a hierarchical chunk structure as increasing durations of pause lengths between written elements. In this research, differences between every pause level within this structure were also found to be significantly different in a single multilevel statistical test. Although the outcomes of the previous experiments had to be carefully qualified, it does

Table 2: Pauses (ms) for various stimulus levels over different stimulus types (rounded to 10 ms).

Experiment	Stimuli	L0	L1	L2	L3	L4
Cheng & Rojas-Anaya (2005)	Number sequences	90	280	440		
Cheng & Rojas-Anaya (2006)	Familiar and jumbled phrases	90	270	400		
Cheng & Rojas-Anaya (2008)	Artificial sentence	90	250	440	600	
Present	Natural language paragraphs	90	270	370	570	1130

appear that the effects found are genuine, because of the consistency with the present experiment.

Regarding the primary aim of this experiment of adding a fifth level to the structure, the results show that when a rehearsed stimulus that possesses five hierarchical levels is written, the TCS, which is based on the pauses between written elements, reflects the ordering of the levels. The stroke level pause lengths are the shortest and the duration increases for each successive increment of level, through letter, word, phrase and sentence level. The increase of the pause with the addition of the fifth sentence level is consistent with the proposal that in graphical production of well rehearsed stimuli each successive chunk level requires specific processing to deal with the particular information associated with that level (Cheng & Rojas-Anaya, 2008).

The direct comparison of the approximate absolute values of the pauses associated with each level for this and the previous experiments reveal some interesting patterns (see Table 2). The three prior experiments noted in this table involved graphical production using the same experimental task methodology: specifically, the writing of sequences from memory after rehearsal with one character in one rectangle. The experiments differ in the important respect that each used a different type of stimulus, as indicated in Table 2. The similarity of the absolute values of the pauses over each level across the different experiments is noteworthy, because it suggests that the same underlying processes are responsible for the pattern of pauses irrespective of the nature of the stimulus. The differences between the pauses on successive levels range between 100 and 200 ms, with mean values of L1-L0=178, L2-L1=145 and L3-L2=180 ms. Taking Newell's (1990) estimate of the time scale for elementary deliberated operations as circa 100 ms, this suggest that there is at least one additional operation occurring when preparing to graphically produce an element that is one level higher in the hierarchy. One such operation will be a process to select the next chunk at a particular level. At the beginning of a new phrase this will involve selection of a phrase, a word, a letter and a stroke. The increase in time suggests that this selection occurs serially and is therefore consistent with the predictions of the hierarchical process model (Cheng & Rojas-Anaya, 2008).

The increase in pause length up to the sentence level from the phrase level is more than three times greater than the increase between any of the other levels. As this is new data from just one experiment, some caution must be taken with its interpretation. It seems to suggest that additional operations that occur at this level do not occur at the levels below. The additional time may be an indication that working memory is fully loaded when complex stimuli comprising multiple sentences with several sub phrases are being processed. Furthermore, the additional time may indicate that retrieval from long-term memory is required as the complete stimuli cannot all be held in working memory despite the rehearsal. As there are approximately ten times as many letters in each of the present stimuli as there were in the stimuli of the previous experiments and as the number of chunks is larger than Miller's magical number 7 ± 2 (Miller, 1956), this is a likely interpretation. In future research, verbal working memory measures, such as the digit span task (Wechsler, 1985) and the listening span task (Daneman & Carpenter, 1980), could be used to gain insight into how pause lengths are related to working memory capacity and the possible involvement of long-term memory. However, it should be noted that as there were 19 pauses on the sentence level at the most, the actual value might be less robust than for the other levels, because outliers have a larger influence with a small number of measurement points. For comparison, there were 33 phrase level, 238 word level, and about 1148 letter level pauses (the number of strokes depended on whether cursive of block letters were used).

Another possible interpretation of this large increase in pause length is that, in addition to retrieving a sentence, inhibiting processes take place to suppress the inclination of writing punctuation. A possibility to overcome this problem is to require participants to write punctuation in a separate rectangle. However, in this case, it is unclear whether the pause between the last letter of for example, the sentence and the period (full stop) or the pause between the period and the first letter of the next sentence, should be taken as an indication of pause length. An alternative for future experiments would be to require participants to write punctuation marks in the same rectangle right after the last letter of the sentence or phrase.

In summary, other processes than selection and retrieval processes might also underpin this pattern of pause lengths. Therefore, empirical and modelling studies are conducted to unravel which processes contribute to the increase in pause length accompanied with each level.

Irrespective of the precise explanation for the increase in duration between each hierarchical level, the results of this experiment reconfirms the claim that there is a temporal signal. This signal may be associated with chunking processes and is a source of high resolution information concerning participants' task performance. With appropriately designed tasks, the TCS could provide valuable evidence to probe the relations among the sub-processes that underpin cognitive phenomena.

A logical next step is to investigate whether a sixth level, the paragraph level, can be added to the hierarchical process model. However, as the demands on working and long-term memory will increase even more, such an experiment has to be designed carefully in order to be feasible.

Acknowledgements

We would like to thank Keith Smith of the University of Sussex for his assistance with the recruitment of participants for this study. We would also like to thank Cora Maas of Utrecht University for her assistance in performing the multilevel analysis.

References

Buschke, H. (1976). Learning is organized by chunking. *Journal of Verbal Learning and Verbal Behavior, 15,* 313-324.

Carroll, D. W. (2004). *Psychology of language.* Belmont, KY: Thomson Wadsworth.

Chase, W., & Simon, H. A. (1973). The mind eye's in chess. In W. Chase (Ed.), *Visual information processing*. New York, N.Y.: Academic press.

Cheng, P. C-H., McFadzean, J., & Copeland, L. (2001). Drawing out the temporal structure of induced perceptual chunks. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society* (pp. 200-205). Mahwah, New Jersey: Lawrence Erlbaum.

Cheng, P. C-H., & Rojas-Anaya, H. (2005). Writing out a temporal signal of chunks: Patterns of pauses reflect the induced structure of written number sequences. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 424-429). Mahwah, NJ: Lawrence Erlbaum.

Cheng, P. C-H., & Rojas-Anaya, H. (2006). A temporal signal reveals chunk structure in the writing of word phrases. In R. Sun, & Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 160-165). Mahwah, NJ: Lawrence Erlbaum Associates.

Cheng, P. C-H., & Rojas-Anaya, H. (2007). Measuring Mathematical Formula Writing Competence: An Application of Graphical Protocol Analysis. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the Twenty Ninth Annual Conference of the Cognitive Science Society* (pp. 869-874). Austin, TX: Cognitive Science Society.

Cheng, P. C-H., & Rojas-Anaya, H. (2008). A Graphical Chunk Production Model: Evaluation Using Graphical Protocol Analysis with Artificial Sentences. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (pp. 1972-1977). Austin, TX: Cognitive Science Society.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Learning and Verbal Behavior, 19,* 450-466.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition, 7,* 149-158.

Hox, J. (2002). *Multilevel analysis: Techniques and applications.* London: Lawrence Erlbaum Associates.

McLean, R. S., & Gregg, L. W. (1967). Effects of induced chunking on temporal aspects of serial recitation. *Journal of Experimental Psychology, 74,* 455-459.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for information processing. *Psychological Review, 63,* 81-97.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Obaidellah, U. H., & Cheng, P. C-H. (2009). Graphical production of complex abstract diagrams: drawing out chunks and schemas. In N. Taatgen & H. v. Rijn (Eds.), *Proceedings of the Thirty-first Annual Conference of the Cognitive Science Society* (pp. 2843-2848). Austin, TX: Cognitive Science Society.

Reitman, J. S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology, 8,* 336-356.

Stavig, G. B., & Gibbons, J. D. (1977). Comparing the mean and the median as measures of centrality. *International Statistical Review / Revue Internationale de Statistique, 45,* 63-70.

van Genuchten, E. (2009). Pause Length Extraction Tool (PLET), Computer Software, University of Sussex. Brighton, UK.

van Genuchten, E., Cheng, P. C-H., Leseman, P. P. M., & Moerland, J. (submitted). Detection of a Working Memory Deficit in Dyslexia: Children Writing from Memory. *Dyslexia.*

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins.